DOCUMENT RESUME

ED 243 935 TM 840 248

AUTHOR Andrich, David
TITLE The Attenuation Paradox of Traditional Test Theory as a Breakdown of Local Independence in Person-Item Response Theory.
PUB DATE Apr 84
NOTE 24p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New Orleans, LA, April 24-26, 1984).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Latent Trait Theory; Testing Problems; Test Interpretation; Test Items; *Test Reliability; *Test Theory
IDENTIFIERS *Attentuation Paradox; *Local Independence (Tests); Unidimensionality (Tests)

ABSTRACT

Both the attenuation paradox of traditional test theory and the assumption of local independence in person-item response theory have caused problems in interpretation. This paper demonstrates that the two are related concepts, and, through this demonstration, both are clarified. It is demonstrated that the breakdown of local independence leads to data which manifest the attenuation paradox. A series of simulation studies show the relationship between the degree of violation of local independence and the effects on the traditional KR-20 index of internal consistency. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED243935

THE ATTENUATION PARADOX OF TRADITIONALTEST THEORY AS A BREAKDOWN OF LOCAL INDEPENDENCE IN PERSON-ITEM RESPONSE THEORY

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

David Andrich

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

DAVID ANDRICH

Department of Education

The University of Western Australia

Nedlands Western Australia 6009

TM 840 248

Paper presented at the

National Council of Measurement in Education

New Orleans, 1984.

The Attenuation Paradox of Traditional Test Theory as a Breakdown of Local Independence in Person-Item Response Theory

DAVID ANDRICH

Department of Education

The University of Western Australia

Nedlands Western Australia 6009

Abstract

Both the attenuation paradox of traditional test theory and the assumption of local independence in person-item response theory have caused problems in interpretation. This paper demonstrates that the two are related concepts, and through this demonstration, both are clarified. It is demonstrated that the breakdown of local independence leads to data which manifest the attenuation paradox. A series of simulation studies show the relationship between the degree of violation of local independence and the effects on the traditional KR.20 index of internal consistency.

Acknowledgements

Discussions on aspects of local independence with Graham Douglas and Pender Pedler have contributd to the clarification of ideas in this paper.

The Attenuation Paradox of Traditional Test Theory as a Breakdown of Local Independence in Person-Item Response Theory

David Andrich

1. Introduction

The attenuation paradox of traditional test theory (TTT) has been readily understood. However, even though it is simple to understand with only a rudimentary knowledge of the ideas of simple and multiple correlation, it does not have a simple resolution. Humphreys (1956), in response to Loevinger (1954), presented an argument designed to demonstrate that the paradox is removed when inter-item correlations are viewed in a particular way.

Humphreys' approach in looking for defects in the argumentation is the correct one. Usually, a paradox reveals some faulty, though subtle, reasoning, either in the assumptions or in the process of deduction. Alternatively, it may require further knowledge or information in a field to remove the paradox. Although substantially directed to the correct resolution, the discussion by Humphreys is not complete and seems overly sophisticated. A paradox, as simply conceived as the attenuation paradox, demands an equally simple resolution. The simple resolution, of course, does not have to be trivial - it may be simple and powerful.

If a paradox itself is not trivial and does not demand new knowledge, it can perhaps be accounted for readily by some clear thinking on its own terms. In scientific work, however, the mutually incompatible though ostensibly correct propositions or conclusions associated with a paradox persist just because it is difficult to resolve it from within its own frame of reference. The paradox can be accepted and researchers working in the area

may learn to cope with it in a practical manner without being unduly bothered by it. This is the case reached with the attenuation paradox of TTT.

Sometimes, the shelving of a paradox can be enhanced simply by invoking a different paradigm in the sense of Kuhn (1970) in which the paradox does not appear. So far, this has been the direction of events with the advent of strong true score theory in testing, or as it will be called in this paper, person-item response theory (PIRT).

In other circumstances the paradox is actually resolved by the application of a different paradigm on the same question. It is this latter approach, with the application of the logic of PIRT, which will be pursued for the resolution of the attenuation paradox of TTT.

2. The Attenuation Paradox

It will be instructive to state the attenuation paradox, which had substantial coverage in the early literature on TTT, in an explicit and direct way. The steps in the case may be listed as follows:

(i) the correlation between one test X and a perfectly reliable criterion test Y, termed the validity r_{XY}, is bounded by the square root of the reliability $r_{XX'}$ of test X, that is, $r_{XY} \leq \sqrt{r_{XX'}}$;

(ii) the result in (i) above implies that the greater the reliability of test X, the greater the possibility of an increased validity;

(iii) one way of increasing the reliability of a test is to increase the inter-correlations among the items of a test;

(iv) the inter-correlations between two items are increased as the responses of the individuals on the two items correspond. In the limit, when all the persons obtain either both items right or both wrong, the correlation will be 1.0;

(v) if the limit of (iv) above is reached for all pairs of items, then the maximum reliability is reached when the sample of persons is dichotomized. In that case, only one of the items is sufficient and the rest are redundant;

(vi) but generally, the correlation between a single item and a criterion is less than the correlation between a test composed of many items measuring the same trait and the same criterion; therefore

(vii) the validity of the test must begin to decrease at some point as the inter-item correlations and the reliability continue to increase.

In summary, the reliability of a test may be artificially high in its indication of the limits on validity.

3. Local Independence

With the advent of PIRT, a paradigm in which the attenuation paradox does not appear, one may wonder why the concern with the attenuation paradox. One is that the resolution of a paradox produces an increase in understanding of a given domain of knowledge and is therefore aesthetically pleasing in its own right. The second is that if it can be resolved by using PIRT, then this resolution gives added impetus for the application of PIRT rather than TTT. The third is that the paradox seems to be resolved, not only by considering

PIRT, but in particular, by viewing the attenuation paradox as a breakdown of the universal assumption of local independence in PIRT. Therefore, in resurrecting the attenuation paradox, it may be possible to clarify the assumption of local independence which is not without its own confusing aspects.

The concept of local independence is simple in some ways, yet, less simple in some of its implications. One of its implications has often been used to illustrate the meaning of local independence, even though this often adds further confusion. In fact, it is this implication which promotes the term "local" which is essentially redundant. These points will be elaborated after the assumption of local independence is defined.

Suppose a set of items have a fixed level of difficulty and that the probability of success of person n to item i is written as

$$p_{xni} = p\{x_{ni} | \beta_n, \delta_i\} \tag{1}$$

where x_{ni} is 1 or 0 according to whether or not the item is responded to correctly or otherwise. The assumption of <u>statistical</u> or <u>stochastic independence</u> among responses holds if, for the total vector of responses, (x_{ni}), i = 1, I,

$$p\{(x_{ni}) | \beta_n, (\delta_i)\} = (p_{xn_1})(p_{xn_2})(p_{xn_3})\ldots(p_{xnI}) \ . \tag{2}$$

This is a statistical definition of independence. Further, if there are N persons, the probability of the matrix of responses $((x_{ni}))$ of the N persons to the I items is given by

$$p\{((x_{ni}))|(\beta_n),(\delta_i)\} = \prod_{ni} \prod p_{xni} . \tag{3}$$

In the case of (1), the response vector (x_{ni}) depends on both the parameter β_n of person n and the vector of the difficulties of the items

(δ_1). In specifying equation (2), the term "local" is redundant - the responses according to the model, which includes the specification of the parameters, are statistically independent according to (2). Interestingly, Rasch (1980) did not preface the assumption of independence with the term "local".

Notice that according to (3), the independence generalises to the case of more than one person responding to each item. Because the assumption of independence across persons for a given item is easier to envisage being satisfied, the focus of discussion is generally an equation (2). As will be shown, however, the assumption is not so simple that independence across persons cannot be violated.

Why, then, the term "local"? Apparently, it originates with Lazarsfeld (Lord, 1980:19), and presumably, it serves some purpose. The way it was introduced and the way it is explained by Lord (op.cit.) gives an indication of why it might be in the literature. Paraphrasing Lord, Lazarsfeld's assumption of local independence is defined by

$$p\{x_{ni} = 1 | \beta_n\} = p\{x_{ni} = 1 | \beta_n, (x_{nj}), j = 1, I, i \neq j\}. \qquad (4)$$

That is, the probability of success on item i, given the ability of person n, is the same as the probability of success of the person on item i given this ability and given the performance on all other items. If the probability of a correct response to item i by person n is the same unconditionally and on condition of the responses to the other items, then the response to item i is independent of the responses to the other items. From (4), equation (2) follows except that in (2) the item parameters also are made explicit.

At this point, there still seems to be little need for the term "local", unless it is to signify that "local" implies a single person. But that is an unusual use of "local" - it might be more consistent with existing terminology in the statistical literature to say that there is a "within-person" independence among responses to items.

There are perhaps two related aspects to the above formulation which give rise to the term "local". First is the fact that one cannot test the assumption of local independence with the responses of only one person. Therefore, the responses of many persons must be considered and cast into some appropriate form. This is possible if one recognises that the responses are governed only by the one ability of the person and the item parameters, and that the theoretical response probabilities of two persons with the same ability are identical. Further, and by definition, the response to one item does not affect the response to any other item. As a consequence, the correlation of the responses to two items is 0.0. For example, suppose 100 persons of the same ability responded to two items, one with a probability of failure equal to 0.40 for those persons and the other of 0.80. Then the theoretical distribution of successes and failures would be as follows:

		Item 2		
	Score	0	1	
Item 1	1	48	12	60
	0	32	8	40
		80	20	100

On the other hand, if the 100 persons had varying abilities, then those persons with superior abilities would tend to get both items right and those with inferior abilities would tend to get both wrong. Accordingly, a non zero correlation is created between the responses to the two items. Thus if one localises the ability, the inter-item correlations are 0.0.

The second aspect concerns a conceptualisation of the abilities of a

sample of persons on the ability continuum as a distribution. In that case, one can focus on any point of the continuum and claim that for persons at that point, the responses have a correlation of 0.0. Again, the need appears to locate or localise the abilities to justify a correlation of 0.0 between items.

Of course, in practice, one does not know the abilities, so that one cannot actually group persons by abilities - at best, one can group by some estimate of abilities. Though the value of item-correlations is mentioned routinely in association with local independence, the calculation of correlations to check on the assumption is rarely made. Instead, other procedures which will be mentioned later are used.

Thus if independence prevails, then the inter-item correlation for a pair of items for persons of the same ability is 0.0. For persons of different ability, the correlation is not 0.0. However, even though the correlation is not 0.0, independence is assumed to prevail. This asymmetrical relationship between the generality of independence and the localisation of a 0.0 inter-item correlation may contribute to misunderstandings of the concept of independence. If the term "local" needs to be used at all, then perhaps it could be used to designate that the local correlation is 0.0 - the point being that independence is not only local, but also general.

There are two related ideas captured by the assumption of independence. One is that there is no learning or any other activity operating between responses to items so that they affect each other. Given the definition of independence, this implication seems straightforward. The second is that only the one ability, β_n, governs all the responses and that one can conceive of persons all of whom have the same ability. Thus, if only one parameter is specified for each person, it is assumed that unidimensionality holds (Lumsden, 1976). If unidimensionaltiy is indicated, then so too is independence. Issues of unidimensionalty in general, and as a consequence of independence, seem to be more troublesome than they might be in PIRT.

The key to the clarification of the concepts of both independence and unidimensionality is an appreciation that they are relative, and not absolute, concepts. Much of the literature, influenced by intelligence testing which has concerned itself with isolating mental traits, gives the impression that unidimensional tests capture some simple "pure" trait. Unidimensionality can be attained by tests which capture many traits. Indeed, every human performance involves many traits. The main issue is whether or not the items of the test capture the same complex of traits (Lumsden, 1961).

4. Violating Independence

One way of appreciating a particular concept or assumption in any model is to construe a way of violating it. Here, data are constructed to violate the assumption of independence and are then analysed in a way which bears directly on the attenuation paradox. Violating independence should create a correlation among item responses for persons of the same ability, and should therefore exaggerate the correlation between items. But an aspect of the attenuation paradox is that the intercorrelations among items might be too high. Therefore violating independence in a way which increases inter-item correlations might be expected to manifest the attenuation paradox.

Simulations of responses of 250 persons to items were used to show the relationships. First, a set of 21 items with mean difficulty 0.0 were constructed and responses of persons with mean ability 0.0 were obtained according to Rasch's simple logistic model (SLM). The SLM takes the form

$$p\{x_{ni} = 1 | \beta_n, \delta_i\} = \frac{1}{\gamma_{ni}} \exp(\beta_n - \delta_i) \qquad (6)$$

where $\gamma_{ni} = 1 + \exp(\beta_n - \delta_i)$ is a normalising factor. This is the most appropriate of the PIRT models for the study here because it is the only model

in which the total score $r_n = \Sigma_i x_{ni}$ of the person is sufficient for the estimate of the person's ability (Wright, 1977). The use of the unweighted total score to characterise a person corresponds to the procedure used in TTT.

In studying the attenuation paradox, the traditional KR.20 reliability index will need to be computed. Because it depends on both the distribution of abilities and the spread of item difficulties, both of these features were varied. Two distributions, one with $\beta \sim N(0,1)$ and one with $\beta = 0$, $n = 1, N$ were used. The latter was used to provide an extreme example in which the KR.20 index would be expected to be 0.0 when the independence condition was satisfied. The items were grouped into seven sets of three. In some simulations the items were of equal difficulty within sets while in other cases they were of different difficulty. The difficulties between sets were different. The case of equal difficulty in subsets was used to demonstrate the appropriateness of some aspects of the estimation procedure and evidence of a proper accounting of the dependence.

Finally, the same conditions were replicated with independence violated within six of the seven subsets of items. It was violated by having the response of a person on the second and third items of the subset depend on the responses to the first and second items respectively. This was constructed in the following manner. If the response to the first item was 1, then the ability of the person was increased by a constant C in responding to the second item. If the response to the first item was 0, then the ability of the person was decreased by the same constant C in responding to the second item. The same relationship was constructed between tems 2 and 3 within each set. Clearly, the concept of independence specified by Lazarsfeld was violated in that the responses to some items did depend on responses to other items. Two levels of dependency were constructed, one with C = 0.75 and one with C = 1.50.

The conditions of simulation are summarised in Table 1 which also shows the relevant results of the analyses.

5. Analyses

As exemplified in the discussion above, the case of the attentuation paradox is highlighted most readily by considering the inter-item correlations and their effect on the reliability. Accordingly, the most pertinent index of reliability is the Kuder-Richardson formula 20 (KR.20). This index was therefore calculated for all data sets generated. Besides calculating the KR.20 index, the data were analysed according to two models of PIRT. One model was the SLM according to which data without a violation of independence were generated. The other was a version of the Rasch rating model (Andrich, 1978; Masters, 1982) which takes the form

$$P_{xnj} = \frac{1}{\gamma_{nj}} \exp\left[x(\beta_n - \delta_j) - \sum_{k=1}^{m} \tau_{kj}\right], \qquad (7)$$

where $x \in \{0, 1, \ldots m\}$ is the rating or score assigned to question j,

τ_{kj}, $k = 1, m$ are m thresholds designating boundaries between successive categories for each question,

and $\gamma_{nj} = \sum_{y=0}^{m} \exp\left[y(\beta_n - \delta_j) - \sum_{k=1}^{m} \tau_{kj}\right]$ is a normalising factor.

The appropriate form of the model for the present analysis involves a reparameterisation of the thresholds in (7). Briefly, if θ_j is the half distance between the successive thresholds, that is, $2\theta_j = \tau_{kj} - \tau_{(k-1)j}$, then (7) may be summarised as

$$P_{xnj} = \frac{1}{\gamma_{nj}} \exp\left[x(\beta_n - \delta_j) + x(m-x)\,\theta_j\right]. \qquad (8)$$

This model will be referred to as the dispersion-location model (DLM).

In the DLM, θ_j characterises the dispersion of the responses (Andrich, 1982a). For the application of this model to the data simulated, the dichotomously scored items in the sets of three were treated as a single entity with the sum of the scores on the dichotomously scored items representing the score on the set of three.

The application of the DLM to such data meant that θ captured two characteristics of the items. First, it incorporated the spread of difficulties of the items in each subset. In particular, the greater the spread of difficulties, the greater the likelihood of more central or regressed total scores; accordingly, the greater the value of θ. Second, it incorporated the degree of dependence among the items in a subset. In particular, the greater the dependence, the greater the spread of total scores; accordingly, the smaller the value of θ. The greater spread follows from dependence because success on an item in a subset increases the chances of success on the subsequent item, and failure on an item increases the chances of failure on a subsequent item. Thus the effects of the spread of difficulties and effect of dependence of items within subsets work against each other. In the case where independence is satisfied and where the items are of equal difficulties, the total score on a subset follows the binomial distribution. With three items per subset, implying a maximum score of three and therefore three thresholds, the theoretical value of θ is 0.55 (Andrich, 1984).

In order to relate the KR.20 index of reliability to the analyses based on the SLM and the DLM, it is instructive to note that the KR.20 index can be obtained from the perspective of a standard analysis of variance with items treated as one factor and persons as the other factor (Hoyt, 1941). Let the observed total score, $\sum_{i=1}^{I} x_{ni}$, where $x_{ni} \in \{0, 1\}$, of person n on the test be denoted by Y_n and let the "among-persons" variance be denoted by V[Y]. Further, let the residual variance, after the "among-items" variance is

accounted for be denoted by V[E]. Then the reliability index, which is identical to the KR.20 and therefore may be denoted as such, is given by

$$KR.20 = \frac{V[Y] - V[E]}{V[Y]} = 1 - \frac{V[E]}{V[Y]} .$$

Now in the analyses according to the SLM and the DLM an estimate $\hat{\beta}_n$ for each person and a variance of this estimate $V[\hat{\beta}_n]$ is provided. Accordingly, the observed "among-persons" variance of these estimates, $V[\hat{\beta}]$, may be readily calculated. Further, $V[\hat{\beta}_n]$ may be interpreted as the within-person error variance and its average across persons, $\bar{V}[\hat{\beta}_n]$, readily obtained. Then an index of the analysis of variance form of the KR.20 above is given by

$$r_\beta = \frac{V[\hat{\beta}] - \bar{V}[\hat{\beta}_n]}{V[\hat{\beta}]} = 1 \frac{\bar{V}[\hat{\beta}_n]}{V[\hat{\beta}]} .$$

The value of r_β clearly is conceptually analogous to the KR.20 index, and indeed its values are also similar to those of the KR.20. (Andrich, 1982b). The index r_β has been termed the person-separation index or more simply the separation index.

Both the KR.20 and the r_β are computed where appropriate, but the advantage of this index over the KR.20 index is that it can be computed from analyses according to either the SLM or the DLM which accounts for the dependence among responses. Thus the effect on the reliability can be readily interpreted through r_β .

6. Results

The results of the analyses are summaried in Table 1 for the persons with abilities distributed as N(0,1) and in Table 2 for all $\beta = 0$.

INSERT TABLE 1 HERE

INSERT TABLE 2 HERE

The Table 1 data are considered first, and particularly the set of output shown in the top left hand corner of the Table. This output pertains to the case where the items are of equal difficulty within the subsets and where independence is satisfied. Therefore, we can anticipate certain results and would expect and require that the figures pertaining to the same index would be similar in analyses according to the SLM, DLM, and the traditional analysis. Clearly, the traditional KR.20 and the r_β indices are identical. (These values are always similar but not necessarily identical.) Further in the case where the items are of the same difficulty within subsets and independence is satisfied, the value of $\theta = 0.54$ which captures the spread of the items within the subsets of three, is very close to the theoretical value of 0.55. The results of this first simulation are used as a frame of reference for considering the results of the other analyses, and the values obtained give confidence in using them in this way.

The next set of figures, to the right of the first set and corresponding to C = 0.75, show marginal differences from the first set, but in the way which is expected given the rationale of the attenuation paradox and the violation of independence in the data analysed. First, the KR.20 and the r_β (under the SLM) indices are greater than in the first set where independence is satisfied. Thus the increase in the inter-correlation among items in subsets produced by the violation of independence increases the traditional reliability. Second, the r_β (DLM) is smaller than the r_β (SLM) and KR.20 of the first set. Thus, when there is dependence and this dependence is properly accounted for, the reliability index is reduced as would be expected, rather than increased. Using the second index to

characterise reliability therefore shows that the upper bound on validity is reduced when items are correlated "too" highly.

Finally, the θ value for the subsets needs to be examined. For the six subsets for those items which exhibit dependence, the value $\hat{\theta} = 0.05$ is smaller than in the first set of figures, $\hat{\theta} = 0.54$. This is as expected and shows that θ is affected by the dependence. For the final set, the value $\hat{\theta} = 0.77$ is greater than $\theta = 0.54$. This seems to be a consistent feature and indicates that the estimate of the parameters for that subset is affected by the estimates of values of the other subsets. Such an effect is surprising and no explanation is suggested.

The third set of figures, to the right of the second set, show a similar trend in relation to the first as do the second with the first. One might have expected even greater KR.20 and r_β (SLM) values with the increased dependence factor. The values obtained suggest that some high threshold value has been reached where much greater dependencies would have to be created before their effects appear in such indices. The value of $\hat{\theta} = -0.36$ is even smaller than $\theta = 0.05$ indicating that the greater dependence is manifested in the θ value. The value of r_β (DLM) = 0.68 is smaller than the r_β = 0.76 value, indicating that the dependence is manifested in the DLM. Thus the effect of greater dependence is to reduce further the reliability and therefore, when calculated to account for dependence, to reduce the upper bound on validity.

The second row of figures in Table 1 are modified by small changes in difficulty of items within subsets. It would seem that these differences have virtually no effect on the resultant data when compared to those of the previous analysis. This perhaps is not surprising considering that the dependence effect was generated by constant changes of ability of 0.75 and 1.50, while the differences in item difficulties were only 0.20. Thus the effects of dependence have swamped the effects of differing difficulties.

The results of Table 2 pertain to the unusual case of a zero true variance of abilities. However, this situation can highlight the nature of certain features unhindered by the effects of the original variation in person abilities. The first set of figures in this Table are predictable from the theory. First, the KR.20 and both r_β indices should have a value close to 0.0. This is indeed the case. Second, the value of θ should be close to 0.54, and again, this is the case. Thus the figures which provide a frame of reference are as expected.

The next set of figures show clearly the effect of dependence. The KR.20 and r_β (SLM) indices are of the order of 0.30, much greater than 0.0. However, the r_β (DLM) index, which accounts for the dependence, remains in the order of 0.0. The value of $\hat{\theta}$ = 0.06 now is much smaller than 0.54. The third set of figures reflects the effect of the greater dependence than the previous set. Now the KR.20 and the r_β (SLM) are of the order of 0.55 while r_β (DLM) has the small value of 0.13. The value of θ = - 0.38 now is even smaller, as might be expected.

The results in the lower half of Table 2, which pertain to the case of items of slightly different difficultly within subsets, provide similar results to the top half of the Table.

7. Discussion

The discussion focusses on lessons to be learnt from a control of the attenuation paradox of traditional test theory and the violation of independence in PIRT by the same mechanism.

First, it is evident from the results of the simulations that the mechanism chosen does produce the required effects. By definition, the mechanism violated independence, and the effect was to increase the KR.20

reliability value. However, when the corresponding reliability was computed by a method which accounted for the dependence among responses, then the reliabilty was shown to be decreased, as might be expected. These results indicate that the KR.20 value was inflated as far as the reliable separation among persons, obtained by computing the total score on the test for each person, is concerned. Therefore it seems established satisfactorily that it is the violation of independence of responses within subsets of items with respect to each person which creates an inflated KR.20 index. It is this inflation which creates the impression of a possible increase in validity when in fact the maximum validity is decreased.

The second point concerns the way the violation of independence was effected. It was effected by modifying a parameter of the model as a function of a response. Thus independence can be violated through a parametric effect, showing that independence and the structure of the parameters are related. This relationship suggests that the manipulation of parameters in ways other than that described above could also violate independence. For example, rather than simply modifying parameters of item subsets according to responses to other items, one might modify the parameters of different item subsets according to some other principle. One such principle is to modify the parameter of a person simply by different amounts for different subtests.

A third point, then, arises from the previous one. This is that when the person parameters are modified differently with respect to different subsets of items, then the unidimensionality condition is also violated. That is, in violating independence, one also violates unidimensionality. Conversely, if one were to focus on the unidimensionality condition and then violate it, one would also violate independence. Perceived in either way, it is evident that introducing effects which disturb the unidimensionality condition contaminates the quality of person separation on the main dimension. A relevant question to ask, of course, is: which is the main dimension? In the SLM and

in traditional test theory, where the total score for each person is obtained simply by adding the scores across all items, the main dimension is the first principal component in the data. It follows then that to the degree that other components manifest themselves in the data, to that degree the separation of persons on the first component will appear to be increased: however, this increase is artificial because it does not result from variation on the first principal component.

A fourth point, again following from the previous one and from the symmetry in the models, can be highlighted. The independence and unidimensionality condition could be violated if the difficulties of items were modified with respect to subsets of persons. Thus one may modify the abilities of subsets of persons in different ways or the difficulties of subsets of items in different ways and so not accommodate this important assumption. In real data, one might expect such subsets of persons and items. For example, subsets of persons taught in different schools and items pertaining to different identifiable classes of topics are found in the same test. The question is not whether unidimensionality or independence is violated, but the degree to which it is violated and the consequences of this violation. Studies of residuals with respect to PIRT models, particularly the SLM model, and especially across persons, are required for this purpose. Any patterns in the residuals show violations of independence or unidimensionality in one way or another. This approach to studying test data is well covered in Wright and Stone (1979) and Wright and Masters (1982).

It might be instructive to dwell a little on the relative nature of unidimensionality. Suppose one has an omnibus reasoning test in which unidimensionality (and therefore independence) is satisfied. Suppose further that there is only one item on mechanical reasoning. Then the addition of another item on mechanical reasoning would violate the independence assumption with respect to all items. The two items on meachanical reasoning would show

a correlation for all persons of the same ability as estimated from the total score of a test. However, one could have twenty items on mechanical reasoning in a test on their own and have them satisfy the unidimensionality and independence assumptions. The test with only mechanical reasoning items would measure a narrower ability than the first test, but both may satisfy the independence assumption. It is important to appreciate the relative nature of unidimensionality and independence.

REFERENCES

Andrich, D. An extension of the Rasch model for ratings providing both location and dispersion parameters. Psychometrika, 1982a, 47, 105-113.

Andrich, D. A rating formulation for ordered response categories. Psychometrika, 1978, 43, 561-573.

Andrich, D. A latent trait model which takes account of dependencies among responses: Implications for test construction and analysis. In S. Embretson (Ed.). Test design: Contributions from psychology education and psychometrics. Academic Press, New York, 1984.

Andrich, D. An index of person separation in latent trait theory, the traditional KR.20 index, and the Guttman scale response pattern. Education Research and Perspectives, 1982b, 9, 95-104.

Hoyt, C. J. Test reliability estimated by analysis of variance. Psychometrika, 1941, 6, 153-160.

Humphreys, L. G. The normal curve and the attenuation paradox in test theory. Psychological Bulletin, 1956, 53, 472-476.

Kuhn, T. S. The structure of scientific revolutions. Chicago: University of Chicago Press, 1970.

Loevinger, J. The attenuation paradox in test theory. Psychological Bulletin, 51, 493-504, 1954.

Lord, F. M. Applications of item response theory to practical testing problems. Lawrence Erbaum Associates. Hillsdale, New Jersey, 1980.

Lumsden, J. Test Theory. Annual Review of Psychology, 1976, 27, 251-280.

Lumsden, J. A factorial approach to unidimensionality. Australian Journal of Psychology, 1957, 9, 105-111.

Masters, G.N. A Rasch model for partial credit scoring. Psychometrika, 1982, 47, 149-174.

Wright, B.D. and Masters, G.N. Rating Scale Analysis, Chicago, MESA Press, 1982.

Wright, B.D. and Stone, M.H. Best Test Design. Chicago, MESA Press, 1979.

Wright, B.D. Solving measurement problems with the Rasch model. Journal of Educational Measurement, 1977, 14, 97-116.

Table 1

Results of analyses of simulated data with person distribution N(0, 1)

Number of persons = 250.

Items with equal difficulty in subsets

δ :	-1.8	-1.8	-1.8	-1.2	-1.2	-1.2	-0.6	-0.6	-0.6	0.0	0.0	0.0
	0.6	0.6	0.6	1.0	1.0	1.0	1.8	1.8	1.8			

	Independence	Dependence	
	C = 0.00	C = 0.75	C = 1.50
KR.20	0.79	0.83	0.83
(SLM) r_β	0.79	0.83	0.83
(DLM) r_β	0.79	0.76	0.68
Ave. $\hat{\theta}$	0,54	Dependent subsets 0.05	-0.36
		Independent subset 0.77	0.67

Items with unequal difficulties in subsets

δ :	-2.0	-1.8	-1.6	-1.4	-1.2	-1.0	-0.8	-0.6	-0.4	-0.2	0.0	0.2
	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0			

	Independence	Dependence	
	C = 0.00	C = 0.75	C = 1.50
KR.20		0.84	0.83
(SLM) r_β		0.84	0.83
(DLM) r_β		0.77	0.68
Ave. $\hat{\theta}$		Dependent subsets 0.07	-0.36
		Independent subset 0.73	0.64

Table 2

Results of analyses of simulated data with person distribution $\beta_n = 0.0$, $n = 1, N$.

Number of persons = 250.

Items with equal difficulty - as Table 1 in subsets

	Independence		Dependence	
			C = 0.75	C = 1.50
KR.20	0.06		0.31	0.54
(SLM) r_β	0.09		0.32	0.55
(DLM) r_β	0.10		0.04	0.13
Ave $\hat{\theta}$	0.57	Dependent subsets	0.10	-0.39
		Independent subsets	0.64	0.63

Items with unequal difficulty in subsets - as Table 1

	Independence		Dependence	
			C = 0.75	C = 1.50
KR.20			0.32	0.53
(SLM) r_β			0.33	0.54
(DLM) r_β			0.03	0.13
Ave $\hat{\theta}$		Dependent subsets	0.06	-0.38
		Independent subsets	0.64	0.65